

RAW SEQUENCE LISTING  
PATENT APPLICATION US/07/920,519

#18

DATE: 09/22/93  
TIME: 09:25:53

INPUT SET: S532.raw

## SEQUENCE LISTING

ENTERED

## (1) General Information:

(i) APPLICANT: CAPUT, DANIEL  
FERRARA, PASCUAL  
GUILLEMOT, JEAN-CLAUDE  
KAGHAD, MOURAD  
LEGOUX, RICHARD  
LOISON, GERARD  
LARBRE, ELIZABETH  
LUPKER, JOHANNES  
LEPLATOIS, PASCUAL  
SALOME, MARK

(ii) TITLE OF INVENTION: URATE OXIDASE ACTIVITY PROTEIN,  
RECOMBINANT GENE CODING THEREFOR, EXPRESSION VECTOR,  
MICRO-ORGANISMS AND TRANSFORMED CELLS

(iii) NUMBER OF SEQUENCES: 36

## (iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: Foley & Lardner  
(B) STREET: 1800 Diagonal Road, Suite 500  
(C) CITY: Alexandria  
(D) STATE: Virginia  
(E) COUNTRY: USA  
(F) ZIP: 22313-0299

## (v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk  
(B) COMPUTER: IBM PC compatible  
(C) OPERATING SYSTEM: PC-DOS/MS-DOS  
(D) SOFTWARE: PatentIn Release #1.0, Version #1.25

## (vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER: US/07/920,519  
(B) FILING DATE:  
(C) CLASSIFICATION:

## (vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: US/07/659,408  
(B) FILING DATE:

## (viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: BENT, Stephen A.  
(B) REGISTRATION NUMBER: 29,768  
(C) REFERENCE/DOCKET NUMBER: 16781/276 BEDL

## (ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: (703)836-9300

RAW SEQUENCE LISTING  
PATENT APPLICATION US/07/920,519DATE: 09/22/93  
TIME: 09:25:59

INPUT SET: S532.raw

52 (B) TELEFAX: (703)683-4109  
53 (C) TELEX: 899149  
54  
55  
56 (2) INFORMATION FOR SEQ ID NO:1:  
57  
58 (i) SEQUENCE CHARACTERISTICS:  
59 (A) LENGTH: 301 amino acids  
60 (B) TYPE: amino acid  
61 (D) TOPOLOGY: linear  
62  
63 (ii) MOLECULE TYPE: protein  
64  
65 (iii) HYPOTHETICAL: NO  
66  
67 (vi) ORIGINAL SOURCE:  
68 (A) ORGANISM: Aspergillus flavus  
69  
70 (vii) IMMEDIATE SOURCE:  
71 (B) CLONE: Urate oxidase  
72  
73  
74 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:  
75  
76 Ser Ala Val Lys Ala Ala Arg Tyr Gly Lys Asp Asn Val Arg Val Tyr  
77 1 5 10 15  
78  
79 Lys Val His Lys Asp Glu Lys Thr Gly Val Gln Thr Val Tyr Glu Met  
80 20 25 30  
81  
82 Thr Val Cys Val Leu Leu Glu Gly Glu Ile Glu Thr Ser Tyr Thr Lys  
83 35 40 45  
84  
85 Ala Asp Asn Ser Val Ile Val Ala Thr Asp Ser Ile Lys Asn Thr Ile  
86 50 55 60  
87  
88 Tyr Ile Thr Ala Lys Gln Asn Pro Val Thr Pro Pro Glu Leu Phe Gly  
89 65 70 75 80  
90  
91 Ser Ile Leu Gly Thr His Phe Ile Glu Lys Tyr Asn His Ile His Ala  
92 85 90 95  
93  
94 Ala His Val Asn Ile Val Cys His Arg Trp Thr Arg Met Asp Ile Asp  
95 100 105 110  
96  
97 Gly Lys Pro His Pro His Ser Phe Ile Arg Asp Ser Glu Glu Lys Arg  
98 115 120 125  
99  
100 Asn Val Gln Val Asp Val Val Glu Gly Lys Gly Ile Asp Ile Lys Ser  
101 130 135 140  
102

RAW SEQUENCE LISTING  
PATENT APPLICATION US/07/920,519DATE: 09/22/93  
TIME: 09:26:03

INPUT SET: S532.raw

```
103      Ser Leu Ser Gly Leu Thr Val Leu Lys Ser Thr Asn Ser Gln Phe Trp
104      145                      150                      155                      160
105
106      Gly Phe Leu Arg Asp Glu Tyr Thr Thr Leu Lys Glu Thr Trp Asp Arg
107                      165                      170                      175
108
109      Ile Leu Ser Thr Asp Val Asp Ala Thr Trp Gln Trp Lys Asn Phe Ser
110                      180                      185                      190
111
112      Gly Leu Gln Glu Val Arg Ser His Val Pro Lys Phe Asp Ala Thr Trp
113                      195                      200                      205
114
115      Ala Thr Ala Arg Glu Val Thr Leu Lys Thr Phe Ala Glu Asp Asn Ser
116                      210                      215                      220
117
118      Ala Ser Val Gln Ala Thr Met Tyr Lys Met Ala Glu Gln Ile Leu Ala
119      225                      230                      235                      240
120
121      Arg Gln Gln Leu Ile Glu Thr Val Glu Tyr Ser Leu Pro Asn Lys His
122                      245                      250                      255
123
124      Tyr Phe Glu Ile Asp Leu Ser Trp His Lys Gly Leu Gln Asn Thr Gly
125                      260                      265                      270
126
127      Lys Asn Ala Glu Val Phe Ala Pro Gln Ser Asp Pro Asn Gly Leu Ile
128                      275                      280                      285
129
130      Lys Cys Thr Val Gly Arg Ser Ser Leu Lys Ser Lys Leu
131      290                      295                      300
132
```

## (2) INFORMATION FOR SEQ ID NO:2:

```
133
134
135      (i) SEQUENCE CHARACTERISTICS:
136          (A) LENGTH: 302 amino acids
137          (B) TYPE: amino acid
138          (C) STRANDEDNESS: single
139          (D) TOPOLOGY: linear
140
141      (ii) MOLECULE TYPE: protein
142
143      (iii) HYPOTHETICAL: NO
144
145      (vi) ORIGINAL SOURCE:
146          (A) ORGANISM: Aspergillus flavus
147
148      (vii) IMMEDIATE SOURCE:
149          (B) CLONE: Met-Urate oxidase
150
151
152      (xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:
153
```

|     |  |     |     |     |     |     |     |     |     |     |       |     |     |     |     |     |     |
|-----|--|-----|-----|-----|-----|-----|-----|-----|-----|-----|-------|-----|-----|-----|-----|-----|-----|
| 154 |  | Met | Ser | Ala | Val | Lys | Ala | Ala | Arg | Tyr | Gly   | Lys | Asp | Asn | Val | Arg | Val |
| 155 |  | 1   |     |     |     | 5   |     |     |     |     | 10    |     |     |     |     | 15  |     |
| 156 |  |     |     |     |     |     |     |     |     |     |       |     |     |     |     |     |     |
| 157 |  | Tyr | Lys | Val | His | Lys | Asp | Glu | Lys | Thr | Gly   | Val | Gln | Thr | Val | Tyr | Glu |
| 158 |  |     |     |     | 20  |     |     |     |     | 25  |       |     |     |     | 30  |     |     |
| 159 |  |     |     |     |     |     |     |     |     |     |       |     |     |     |     |     |     |
| 160 |  | Met | Thr | Val | Cys | Val | Leu | Leu | Glu | Gly | Glu   | Ile | Glu | Thr | Ser | Tyr | Thr |
| 161 |  |     |     | 35  |     |     |     |     | 40  |     |       |     |     | 45  |     |     |     |
| 162 |  |     |     |     |     |     |     |     |     |     |       |     |     |     |     |     |     |
| 163 |  | Lys | Ala | Asp | Asn | Ser | Val | Ile | Val | Ala | Thr   | Asp | Ser | Ile | Lys | Asn | Thr |
| 164 |  |     | 50  |     |     |     |     | 55  |     |     |       |     | 60  |     |     |     |     |
| 165 |  |     |     |     |     |     |     |     |     |     |       |     |     |     |     |     |     |
| 166 |  | Ile | Tyr | Ile | Thr | Ala | Lys | Gln | Asn | Pro | Val   | Thr | Pro | Pro | Glu | Leu | Phe |
| 167 |  | 65  |     |     |     |     | 70  |     |     |     |       | 75  |     |     |     |     | 80  |
| 168 |  |     |     |     |     |     |     |     |     |     |       |     |     |     |     |     |     |
| 169 |  | Gly | Ser | Ile | Leu | Gly | Thr | His | Phe | Ile | Glu   | Lys | Tyr | Asn | His | Ile | His |
| 170 |  |     |     |     | 85  |     |     |     |     |     | 90    |     |     |     |     | 95  |     |
| 171 |  |     |     |     |     |     |     |     |     |     |       |     |     |     |     |     |     |
| 172 |  | Ala | Ala | His | Val | Asn | Ile | Val | Cys | His | Arg   | Trp | Thr | Arg | Met | Asp | Ile |
| 173 |  |     |     |     | 100 |     |     |     |     | 105 |       |     |     |     | 110 |     |     |
| 174 |  |     |     |     |     |     |     |     |     |     |       |     |     |     |     |     |     |
| 175 |  | Asp | Gly | Lys | Pro | His | Pro | His | Ser | Phe | Ile   | Arg | Asp | Ser | Glu | Glu | Lys |
| 176 |  |     |     | 115 |     |     |     |     | 120 |     |       |     |     | 125 |     |     |     |
| 177 |  |     |     |     |     |     |     |     |     |     |       |     |     |     |     |     |     |
| 178 |  | Arg | Asn | Val | Gln | Val | Asp | Val | Glu | Gly | Lys   | Gly | Ile | Asp | Ile | Lys |     |
| 179 |  |     | 130 |     |     |     |     | 135 |     |     |       |     | 140 |     |     |     |     |
| 180 |  |     |     |     |     |     |     |     |     |     |       |     |     |     |     |     |     |
| 181 |  | Ser | Ser | Leu | Ser | Gly | Leu | Thr | Val | Leu | Lys   | Ser | Thr | Asn | Ser | Gln | Phe |
| 182 |  | 145 |     |     |     |     | 150 |     |     |     |       | 155 |     |     |     |     | 160 |
| 183 |  |     |     |     |     |     |     |     |     |     |       |     |     |     |     |     |     |
| 184 |  | Trp | Gly | Phe | Leu | Arg | Asp | Glu | Tyr | Thr | Thr   | Leu | Lys | Glu | Thr | Trp | Asp |
| 185 |  |     |     |     | 165 |     |     |     |     |     | 170   |     |     |     |     | 175 |     |
| 186 |  |     |     |     |     |     |     |     |     |     |       |     |     |     |     |     |     |
| 187 |  | Arg | Ile | Leu | Ser | Thr | Asp | Val | Asp | Ala | Thr   | Trp | Gln | Trp | Lys | Asn | Phe |
| 188 |  |     |     |     | 180 |     |     |     |     | 185 |       |     |     |     | 190 |     |     |
| 189 |  |     |     |     |     |     |     |     |     |     |       |     |     |     |     |     |     |
| 190 |  | Ser | Gly | Leu | Gln | Glu | Val | Arg | Ser | His | Val   | Pro | Lys | Phe | Asp | Ala | Thr |
| 191 |  |     |     | 195 |     |     |     |     | 200 |     |       |     |     | 205 |     |     |     |
| 192 |  |     |     |     |     |     |     |     |     |     |       |     |     |     |     |     |     |
| 193 |  | Trp | Ala | Thr | Ala | Arg | Glu | Val | Thr | Leu | Lys   | Thr | Phe | Ala | Glu | Asp | Asn |
| 194 |  |     | 210 |     |     |     |     | 215 |     |     |       |     | 220 |     |     |     |     |
| 195 |  |     |     |     |     |     |     |     |     |     |       |     |     |     |     |     |     |
| 196 |  | Ser | Ala | Ser | Val | Gln | Ala | Thr | Met | Tyr | Lys</ |     |     |     |     |     |     |

RAW SEQUENCE LISTING  
PATENT APPLICATION US/07/920,519DATE: 09/22/93  
TIME: 09:26:23

INPUT SET: S532.raw

205 Gly Lys Asn Ala Glu Val Phe Ala Pro Gln Ser Asp Pro Asn Gly Leu  
206 275 280 285  
207  
208 Ile Lys Cys Thr Val Gly Arg Ser Ser Leu Lys Ser Lys Leu  
209 290 295 300  
210

## (2) INFORMATION FOR SEQ ID NO:3:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 906 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA (genomic)

## (vii) IMMEDIATE SOURCE:

- (B) CLONE: Preferred sequence for expression in prokaryotes

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

229 ATGTCTGCGG TAAAGCAGC GCGCTACGGC AAGGACAATG TTCGCGTCTA CAAGGTTTAC 60  
230  
231 AAGGACGAGA AGACCGGTGT CCAGACGGTG TACGAGATGA CCGTCTGTGT GCTTCTGGAG 120  
232  
233 GGTGAGATTG AGACCTCTTA CACCAAGGCC GACAACAGCG TCATTGTCGC AACCGACTCC 180  
234  
235 ATTAAGAACA CCATTTACAT CACCGCCAAG CAGAACCCCG TTAATCCTCC CGAGCTGTTC 240  
236  
237 GGCTCCATCC TGGGCACACA CTTTATTGAG AAGTACAACC ACATCCATGC CGCTCACGTC 300  
238  
239 AACATTGTCT GCCACCGCTG GACCCGGATG GACATTGACG GCAAGCCACA CCCTCACTCC 360  
240  
241 TTCATCCGCG ACAGCGAGGA GAAGCGGAAT GTGCAGGTGG ACGTGGTCGA GGGCAAGGGC 420  
242  
243 ATCGATATCA AGTCGTCTCT GTCCGGCCTG ACCGTGCTGA AGAGCACCAA CTCGCAGTTC 480  
244  
245 TGGGGCTTCC TGCGTGACGA GTACACCACA CTTAAGGAGA CCTGGGACCG TATCCTGAGC 540  
246  
247 ACCGACGTCG ATGCCACTTG GCAGTGGAAG AATTTTCAGT GACTCCAGGA GGTCCGCTCG 600  
248  
249 CACGTGCCTA AGTTGATGC TACCTGGGCC ACTGCTCGCG AGGTCACTCT GAAGACTTTT 660  
250  
251 GCTGAAGATA ACAGTGCCAG CGTGCAGGCC ACTATGTACA AGATGGCAGA GCAAATCCTG 720  
252  
253 GCGCGCCAGC AGCTGATCGA GACTGTGAG TACTCGTTGC CTAACAAGCA CTATTTTGAA 780  
254  
255 ATCGACCTGA GCTGGCACAA GGGCCTCCAA AACACCGGCA AGAACGCCGA GGTCTTCGCT 840

RAW SEQUENCE LISTING  
PATENT APPLICATION US/07/920,519DATE: 09/22/93  
TIME: 09:26:31

INPUT SET: S532.raw

256  
257 CCTCAGTCGG ACCCCAACGG TCTGATCAAG TGTACCGTCG GCCGGTCCTC TCTGAAGTCT 900  
258  
259 AAATTG 906  
260  
261 (2) INFORMATION FOR SEQ ID NO:4:  
262  
263 (i) SEQUENCE CHARACTERISTICS:  
264 (A) LENGTH: 906 base pairs  
265 (B) TYPE: nucleic acid  
266 (C) STRANDEDNESS: single  
267 (D) TOPOLOGY: linear  
268  
269 (ii) MOLECULE TYPE: DNA (genomic)  
270  
271  
272 (vii) IMMEDIATE SOURCE:  
273 (B) CLONE: Preferred sequence for expression in  
274 eukaryotes  
275  
276  
277 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:  
278  
279 ATGTCTGCTG TTAAGGCTGC TAGATACGGT AAGGACAACG TTAGAGTCTA CAAGGTTTAC 60  
280  
281 AAGGACGAGA AGACCGGTGT CCAGACGGTG TACGAGATGA CCGTCTGTGT GCTTCTGGAG 120  
282  
283 GGTGAGATTG AGACCTCTTA CACCAAGGCC GACAACAGCG TCATTGTGCG AACC GACTCC 180  
284  
285 ATTAAGAACA CCATTTACAT CACCGCCAAG CAGAACCCCG TTAATCCTCC CGAGCTGTTC 240  
286  
287 GGCTCCATCC TGGGCACACA CTTTATTGAG AAGTACAACC ACATCCATGC CGCTCACGTC 300  
288  
289 AACATTGTCT GCCACGCTG GACCCGGATG GACATTGACG GCAAGCCACA CCCTCACTCC 360  
290  
291 TTCATCCGCG ACAGCGAGGA GAAGCGGAAT GTGCAGGTGG ACGTGGTCGA GGGCAAGGGC 420  
292  
293 ATCGATATCA AGTCGTCTCT GTCCGGCCTG ACCGTGCTGA AGAGCACCAA CTCGCAGTTC 480  
294  
295 TGGGGCTTCC TGCGTGACGA GTACACCACA CTTAAGGAGA CCTGGGACCG TATCCTGAGC 540  
296  
297 ACCGACGTCG ATGCCACTTG GCAGTGGAAG AATTTTCAAGT GACTCCAGGA GGTCCGCTCG 600  
298  
299 CACGTGCCTA AGTTTCGATGC TACCTGGGCC ACTGCTCGCG AGGTCCTCT GAAGACTTTT 660  
300  
301 GCTGAAGATA ACAGTGCCAG CGTGCAGGCC ACTATGTACA AGATGGCAGA GCAAATCCTG 720  
302  
303 GCGCGCCAGC AGCTGATCGA GACTGTGAG TACTCGTTGC CTAACAAGCA CTATTTTGAA 780  
304  
305 ATCGACCTGA GCTGGCACAA GGGCCTCCAA AACACCGGCA AGAACGCCGA GGTCTTCGCT 840  
306

RAW SEQUENCE LISTING  
PATENT APPLICATION US/07/920,519DATE: 09/22/93  
TIME: 09:26:39

INPUT SET: S532.raw

307 CCTCAGTCGG ACCCCAACGG TCTGATCAAG TGTACCGTCG GCCGGTCCTC TCTGAAGTCT 900  
308  
309 AAATTG 906  
310  
311 (2) INFORMATION FOR SEQ ID NO:5:  
312  
313 (i) SEQUENCE CHARACTERISTICS:  
314 (A) LENGTH: 14 base pairs  
315 (B) TYPE: nucleic acid  
316 (C) STRANDEDNESS: single  
317 (D) TOPOLOGY: linear  
318  
319 (ii) MOLECULE TYPE: DNA (genomic)  
320  
321 (iii) HYPOTHETICAL: NO  
322  
323  
324 (vii) IMMEDIATE SOURCE:  
325 (B) CLONE: Preferred non-translated 5' sequence for  
326 animal cells  
327  
328  
329 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:  
330  
331 AGCTTGCCGC CACT 14  
332  
333 (2) INFORMATION FOR SEQ ID NO:6:  
334  
335 (i) SEQUENCE CHARACTERISTICS:  
336 (A) LENGTH: 906 base pairs  
337 (B) TYPE: nucleic acid  
338 (C) STRANDEDNESS: double  
339 (D) TOPOLOGY: linear  
340  
341 (ii) MOLECULE TYPE: DNA (genomic)  
342  
343 (iii) HYPOTHETICAL: NO  
344  
345  
346 (vii) IMMEDIATE SOURCE:  
347 (B) CLONE: Preferred sequence for expression in animal  
348 cells  
349  
350  
351 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:  
352  
353 ATGTCCGCAG TAAAGCAGC CCGCTACGGC AAGGACAATG TCCGCGTCTA CAAGGTTAC 60  
354  
355 AAGGACGAGA AGACCGGTGT CCAGACGGTG TACGAGATGA CCGTCTGTGT GCTTCTGGAG 120  
356  
357 GGTGAGATTG AGACCTCTTA CACCAAGGCC GACAACAGCG TCATTGTCGC AACCGACTCC 180

RAW SEQUENCE LISTING  
PATENT APPLICATION US/07/920,519DATE: 09/22/93  
TIME: 09:26:47

INPUT SET: S532.raw

358  
359 ATTAAGAACA CCATTTACAT CACCGCCAAG CAGAACCCCG TTACTCCTCC CGAGCTGTTC 240  
360  
361 GGCTCCATCC TGGGCACACA CTTCATTGAG AAGTACAACC ACATCCATGC CGCTCACGTC 300  
362  
363 AACATTGTCT GCCACCGCTG GACCCGGATG GACATTGACG GCAAGCCACA CCCTCACTCC 360  
364  
365 TTCATCCGCG ACAGCGAGGA GAAGCGGAAT GTGCAGGTGG ACGTGGTCGA GGGCAAGGGC 420  
366  
367 ATCGATATCA AGTCGTCTCT GTCCGGCCTG ACCGTGCTGA AGAGCACCAA CTCGCAGTTC 480  
368  
369 TGGGGCTTCC TCGGTGACGA GTACACCACA CTTAAGGAGA CCTGGGACCG TATCCTGAGC 540  
370  
371 ACCGACGTCG ATGCCACTTG GCAGTGGAAG AATTTCAAGT GACTCCAGGA GGTCCGCTCG 600  
372  
373 CACGTGCCTA AGTTCGATGC TACCTGGGCC ACTGCTCGCG AGGTCCTCT GAAGACTTTT 660  
374  
375 GCTGAAGATA ACAGTGCCAG CGTGCAGGCC ACTATGTACA AGATGGCAGA GCAAATCCTG 720  
376  
377 GCGCGCCAGC AGCTGATCGA GACTGTGAG TACTCGTTGC CTAACAAGCA CTATTTCGAA 780  
378  
379 ATCGACCTGA GCTGGCACAA GGGCCTCCAA AACACCGGCA AGAACGCCGA GGTCTTCGCT 840  
380  
381 CCTCAGTCGG ACCCCAACGG TCTGATCAAG TGTACCGTCG GCCGGTCCTC TCTGAAGTCT 900  
382  
383 AAATTG 906  
384

## (2) INFORMATION FOR SEQ ID NO:7:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 23 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA (genomic)

## (iii) HYPOTHETICAL: NO

## (vii) IMMEDIATE SOURCE:

- (B) CLONE: reverse transcription primer

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

GATCCGGGCC CTTTTTTTTT TTT 23

## (2) INFORMATION FOR SEQ ID NO:8:

## (i) SEQUENCE CHARACTERISTICS:

RAW SEQUENCE LISTING  
PATENT APPLICATION US/07/920,519DATE: 09/22/93  
TIME: 09:26:55

INPUT SET: S532.raw

409 (A) LENGTH: 10 amino acids  
410 (B) TYPE: amino acid  
411 (C) STRANDEDNESS: single  
412 (D) TOPOLOGY: linear  
413  
414 (ii) MOLECULE TYPE: peptide  
415  
416 (iii) HYPOTHETICAL: NO  
417  
418  
419 (vii) IMMEDIATE SOURCE:  
420 (B) CLONE: Hydrolysis product T 17  
421  
422  
423 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:  
424  
425 Asn Val Gln Val Asp Val Val Glu Gly Lys  
426 1 5 10  
427  
428 (2) INFORMATION FOR SEQ ID NO:9:  
429  
430 (i) SEQUENCE CHARACTERISTICS:  
431 (A) LENGTH: 8 amino acids  
432 (B) TYPE: amino acid  
433 (C) STRANDEDNESS: single  
434 (D) TOPOLOGY: linear  
435  
436 (ii) MOLECULE TYPE: peptide  
437  
438 (iii) HYPOTHETICAL: NO  
439  
440  
441 (vii) IMMEDIATE SOURCE:  
442 (B) CLONE: Hydrolysis product T 20  
443  
444  
445 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:  
446  
447 Asn Phe Ser Gly Leu Gln Glu Val  
448 1 5  
449  
450 (2) INFORMATION FOR SEQ ID NO:10:  
451  
452 (i) SEQUENCE CHARACTERISTICS:  
453 (A) LENGTH: 6 amino acids  
454 (B) TYPE: amino acid  
455 (C) STRANDEDNESS: single  
456 (D) TOPOLOGY: linear  
457  
458 (ii) MOLECULE TYPE: peptide  
459

RAW SEQUENCE LISTING  
PATENT APPLICATION US/07/920,519DATE: 09/22/93  
TIME: 09:27:03

INPUT SET: S532.raw

460 (iii) HYPOTHETICAL: NO  
461  
462  
463 (vii) IMMEDIATE SOURCE:  
464 (B) CLONE: Hydrolysis product T 23  
465  
466  
467 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:  
468  
469 Phe Asp Ala Thr Trp Ala  
470 1 5  
471  
472 (2) INFORMATION FOR SEQ ID NO:11:  
473  
474 (i) SEQUENCE CHARACTERISTICS:  
475 (A) LENGTH: 8 amino acids  
476 (B) TYPE: amino acid  
477 (C) STRANDEDNESS: single  
478 (D) TOPOLOGY: linear  
479  
480 (ii) MOLECULE TYPE: peptide  
481  
482 (iii) HYPOTHETICAL: NO  
483  
484  
485 (vii) IMMEDIATE SOURCE:  
486 (B) CLONE: Hydrolysis product T 27  
487  
488  
489 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:  
490  
491 His Tyr Phe Glu Ile Asp Leu Ser  
492 1 5  
493  
494 (2) INFORMATION FOR SEQ ID NO:12:  
495  
496 (i) SEQUENCE CHARACTERISTICS:  
497 (A) LENGTH: 13 amino acids  
498 (B) TYPE: amino acid  
499 (C) STRANDEDNESS: single  
500 (D) TOPOLOGY: linear  
501  
502 (ii) MOLECULE TYPE: peptide  
503  
504 (iii) HYPOTHETICAL: NO  
505  
506  
507 (vii) IMMEDIATE SOURCE:  
508 (B) CLONE: Hydrolysis product T 28  
509  
510

RAW SEQUENCE LISTING  
PATENT APPLICATION US/07/920,519DATE: 09/22/93  
TIME: 09:27:11

INPUT SET: S532.raw

511 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:  
512  
513 Ile Leu Ser Thr Asp Val Asp Ala Thr Trp Gln Trp Lys  
514 1 5 10  
515

516 (2) INFORMATION FOR SEQ ID NO:13:  
517

518 (i) SEQUENCE CHARACTERISTICS:  
519 (A) LENGTH: 11 amino acids  
520 (B) TYPE: amino acid  
521 (C) STRANDEDNESS: single  
522 (D) TOPOLOGY: linear  
523

524 (ii) MOLECULE TYPE: peptide  
525

526 (iii) HYPOTHETICAL: NO  
527

528  
529 (vii) IMMEDIATE SOURCE:  
530 (B) CLONE: Hydrolysis product T 29  
531

532  
533 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:  
534

535 His Tyr Phe Glu Ile Asp Leu Ser Trp His Lys  
536 1 5 10  
537

538 (2) INFORMATION FOR SEQ ID NO:14:  
539

540 (i) SEQUENCE CHARACTERISTICS:  
541 (A) LENGTH: 11 amino acids  
542 (B) TYPE: amino acid  
543 (C) STRANDEDNESS: single  
544 (D) TOPOLOGY: linear  
545

546 (ii) MOLECULE TYPE: peptide  
547

548 (iii) HYPOTHETICAL: NO  
549

550  
551 (vii) IMMEDIATE SOURCE:  
552 (B) CLONE: Hydrolysis product T 31  
553

554  
555 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:  
556

557 Ser Thr Asn Ser Gln Phe Trp Gly Phe Leu Arg  
558 1 5 10  
559

560 (2) INFORMATION FOR SEQ ID NO:15:  
561

RAW SEQUENCE LISTING  
PATENT APPLICATION US/07/920,519DATE: 09/22/93  
TIME: 09:27:19

INPUT SET: S532.raw

562 (i) SEQUENCE CHARACTERISTICS:  
563 (A) LENGTH: 16 amino acids  
564 (B) TYPE: amino acid  
565 (C) STRANDEDNESS: single  
566 (D) TOPOLOGY: linear  
567  
568 (ii) MOLECULE TYPE: peptide  
569  
570 (iii) HYPOTHETICAL: NO  
571  
572  
573 (vii) IMMEDIATE SOURCE:  
574 (B) CLONE: Hydrolysis product T 32  
575  
576  
577 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:  
578  
579 Gln Asn Pro Val Thr Pro Pro Glu Leu Phe Gly Ser Ile Leu Gly Thr  
580 1 5 10 15  
581  
582  
583 (2) INFORMATION FOR SEQ ID NO:16:  
584  
585 (i) SEQUENCE CHARACTERISTICS:  
586 (A) LENGTH: 16 amino acids  
587 (B) TYPE: amino acid  
588 (C) STRANDEDNESS: single  
589 (D) TOPOLOGY: linear  
590  
591 (ii) MOLECULE TYPE: peptide  
592  
593 (iii) HYPOTHETICAL: NO  
594  
595  
596 (vii) IMMEDIATE SOURCE:  
597 (B) CLONE: Hydrolysis product T 33  
598  
599  
600 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:  
601  
602 Gln Asn Pro Val Thr Pro Pro Glu Leu Phe Gly Ser Ile Leu Gly Thr  
603 1 5 10 15  
604  
605  
606 (2) INFORMATION FOR SEQ ID NO:17:  
607  
608 (i) SEQUENCE CHARACTERISTICS:  
609 (A) LENGTH: 17 amino acids  
610 (B) TYPE: amino acid  
611 (C) STRANDEDNESS: single  
612 (D) TOPOLOGY: linear

RAW SEQUENCE LISTING  
PATENT APPLICATION US/07/920,519DATE: 09/22/93  
TIME: 09:27:27

INPUT SET: S532.raw

613  
614 (ii) MOLECULE TYPE: peptide  
615  
616 (iii) HYPOTHETICAL: NO  
617  
618  
619 (vii) IMMEDIATE SOURCE:  
620 (B) CLONE: Hydrolysis product V 1  
621  
622  
623 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:  
624  
625 Tyr Ser Leu Pro Asn Lys His Tyr Phe Glu Ile Asp Leu Ser Trp His  
626 1 5 10 15  
627  
628 Lys  
629  
630  
631 (2) INFORMATION FOR SEQ ID NO:18:  
632  
633 (i) SEQUENCE CHARACTERISTICS:  
634 (A) LENGTH: 16 amino acids  
635 (B) TYPE: amino acid  
636 (C) STRANDEDNESS: single  
637 (D) TOPOLOGY: linear  
638  
639 (ii) MOLECULE TYPE: peptide  
640  
641 (iii) HYPOTHETICAL: NO  
642  
643  
644 (vii) IMMEDIATE SOURCE:  
645 (B) CLONE: Hydrolysis product V 2  
646  
647  
648 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:  
649  
650 Val Thr Leu Lys Thr Phe Ala Glu Asp Asn Ser Ala Ser Val Gln Ala  
651 1 5 10 15  
652  
653  
654 (2) INFORMATION FOR SEQ ID NO:19:  
655  
656 (i) SEQUENCE CHARACTERISTICS:  
657 (A) LENGTH: 24 amino acids  
658 (B) TYPE: amino acid  
659 (C) STRANDEDNESS: single  
660 (D) TOPOLOGY: linear  
661  
662 (ii) MOLECULE TYPE: peptide  
663

RAW SEQUENCE LISTING  
PATENT APPLICATION US/07/920,519DATE: 09/22/93  
TIME: 09:27:35

INPUT SET: S532.raw

664 (iii) HYPOTHETICAL: NO  
665  
666  
667 (vii) IMMEDIATE SOURCE:  
668 (B) CLONE: Hydrolysis product V 3  
669  
670  
671 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:  
672  
673 Thr Ser Tyr Thr Lys Ala Asp Asn Ser Val Ile Val Asp Thr Asp Ser  
674 1 5 10 15  
675  
676 Ile Lys Asn Thr Ile Tyr Ile Thr  
677 20  
678  
679 (2) INFORMATION FOR SEQ ID NO:20:  
680  
681 (i) SEQUENCE CHARACTERISTICS:  
682 (A) LENGTH: 28 amino acids  
683 (B) TYPE: amino acid  
684 (C) STRANDEDNESS: single  
685 (D) TOPOLOGY: linear  
686  
687 (ii) MOLECULE TYPE: peptide  
688  
689 (iii) HYPOTHETICAL: NO  
690  
691  
692 (vii) IMMEDIATE SOURCE:  
693 (B) CLONE: Hydrolysis product V 5  
694  
695  
696 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:  
697  
698 Gly Lys Gly Ile Asp Ile Lys Ser Ser Leu Ser Gly Leu Thr Val Leu  
699 1 5 10 15  
700  
701 Lys Ser Thr Asn Ser Gln Phe Trp Gly Phe Leu Arg  
702 20 25  
703  
704 (2) INFORMATION FOR SEQ ID NO:21:  
705  
706 (i) SEQUENCE CHARACTERISTICS:  
707 (A) LENGTH: 17 amino acids  
708 (B) TYPE: amino acid  
709 (C) STRANDEDNESS: single  
710 (D) TOPOLOGY: linear  
711  
712 (ii) MOLECULE TYPE: peptide  
713  
714 (iii) HYPOTHETICAL: NO

RAW SEQUENCE LISTING  
PATENT APPLICATION US/07/920,519DATE: 09/22/93  
TIME: 09:27:43

INPUT SET: S532.raw

715  
716  
717 (vii) IMMEDIATE SOURCE:  
718 (B) CLONE: Hydolysis product V 6  
719  
720  
721 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:  
722  
723 Gly Lys Gly Ile Asp Ile Lys Ser Ser Leu Ser Gly Leu Thr Val Leu  
724 1 5 10 15  
725  
726 Lys  
727  
728  
729 (2) INFORMATION FOR SEQ ID NO:22:  
730  
731 (i) SEQUENCE CHARACTERISTICS:  
732 (A) LENGTH: 1236 base pairs  
733 (B) TYPE: nucleic acid  
734 (C) STRANDEDNESS: single  
735 (D) TOPOLOGY: linear  
736  
737 (ii) MOLECULE TYPE: DNA (genomic)  
738  
739 (iii) HYPOTHETICAL: NO  
740  
741  
742 (vii) IMMEDIATE SOURCE:  
743 (B) CLONE: Fragment 3  
744  
745  
746 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:  
747  
748 GATCCGCGGA AGCATAAAGT GTAAAGCCTG GGGTGCCTAA TGAGTGAGCT AACTTACATT 60  
749  
750 AATTGCGTTG CGCTCACTGC CCGCTTTCCA GTCGGGAAAC CTGTCGTGCC AGCTGCATTA 120  
751  
752 ATGAATCGGC CAACGCGCGG GGAGAGGCGG TTTGCGTATT GGGCGCCAGG GTGGTTTTTC 180  
753  
754 TTTTCACCAG TGAGACGGGC AACAGCTGAT TGCCCTTCAC CGCCTGGCCC TGAGAGAGTT 240  
755  
756 GCAGCAAGCG GTCCACGCTG GTTTGCCCCA CCACCCGAAA ATCCTGTTTG ATGGTGGTTA 300  
757  
758 ACGGCGGGAT ATAACATGAG CTGTCTTCGG TATCGTCGTA TCCCACTACC GAGATATCCG 360  
759  
760 CACCAACGCG CAGCCCGGAC TCGGTAATGG CGCGCATTGC GCCCAGCGCC ATCTGATCGT 420  
761  
762 TGGCAACCAG CATCGCAGTG GGAACGATGC CCTCATTCAG CATTTGCATG GTTTGTTGAA 480  
763  
764 AACCGGACAT GGCACCTCCAG TCGCCTTCCC GTTCCGCTAT CGGCTGAATT TGATTGCGAG 540  
765

RAW SEQUENCE LISTING  
PATENT APPLICATION US/07/920,519DATE: 09/22/93  
TIME: 09:27:51

INPUT SET: S532.raw

766 TGAGATATTT ATGCCAGCCA GCCAGACGCA GACGCGCCGA GACAGAACTT AATGGGCCCCG 600  
767  
768 CTAACAGCGC GATTTGCTGG TGACCCAATG CGACCAGATG CTCCACGCCC AGTCGCGTAC 660  
769  
770 CGTCTTCATG GGAGAAAATA ATACTGTTGA TGGGTGTCTG GTCAGAGACA TCAAGAAATA 720  
771  
772 ACGCCGGAAC ATTAGTGCAG GCAGCTTCCA CAGCAATGGC ATCCTGGTCA TCCAGCGGAT 780  
773  
774 AGTTAATGAT CAGCCCACTG ACGCGTTGCG CGAGAAGATT GTGCACCGCC GCTTTACAGG 840  
775  
776 CTTGACGCC GCTTCGTTCT ACCATCGACA CCACCACGCT GGCACCCAGT TGATCGGCGC 900  
777  
778 GAGATTTAAT CGCCGCGACA ATTTGCGACG GCGCGTGCAG GGCCAGACTG GAGGTGGCAA 960  
779  
780 CGCCAATCAG CAACGACTGT TTGCCCGCCA GTTGTGTGTC CACGCGGTTG GGAATGTAAT 1020  
781  
782 TCAGCTCCGC CATCGCCGCT TCCACTTTTT CCCGCGTTTT CGCAGAAACG TGGCTGGCCT 1080  
783  
784 GGTTCACCAC GCGGGAAACG GTCTGATAAC AGACACCGGC ATACTCTGCG ACATCGTATA 1140  
785  
786 ACGTTACTGG TTTCACATTC ACCACCCTGA ATTGACTCTC TTCCGGGCGC TATCATGCCA 1200  
787  
788 TACCGCGAAA GGTTTTGCGC CATTCGATGG TGTCCG 1236  
789

## (2) INFORMATION FOR SEQ ID NO:23:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 321 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA (genomic)

## (iii) HYPOTHETICAL: NO

## (vii) IMMEDIATE SOURCE:

- (B) CLONE: Fragment 4

## (ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 107..316
- (D) OTHER INFORMATION: /product= "regulatory signal + aa  
1-44 human growth hormone precursor"

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

815 TCGAGCTGAC TGACCTGTTG CTTATATTAC ATCGATAGCG TATAATGTGT GGAATTGTGA 60  
816

# RAW SEQUENCE LISTING PATENT APPLICATION US/07/920,519

DATE: 09/22/93  
TIME: 09:27:59

INPUT SET: S532.raw

```

817 GCGATAACAA TTTCACACAG TTTAACTTTA AGAAGGAGAT ATACAT ATG GCT ACC      115
818                                     Met Ala Thr
819                                     1
820
821 GGA TCC CGG ACT AGT CTG CTC CTG GCT TTT GGC CTG CTC TGC CTG CCC      163
822 Gly Ser Arg Thr Ser Leu Leu Leu Ala Phe Gly Leu Leu Cys Leu Pro
823      5                      10                      15
824
825 TGG CTT CAA GAG GGC AGT GCC TTC CCA ACC ATT CCC TTA TCT AGA CTT      211
826 Trp Leu Gln Glu Gly Ser Ala Phe Pro Thr Ile Pro Leu Ser Arg Leu
827      20                      25                      30                      35
828
829 TTT GAC AAC GCT ATG CTC CGC GCC CAT CGT CTG CAC CAG CTG GCC TTT      259
830 Phe Asp Asn Ala Met Leu Arg Ala His Arg Leu His Gln Leu Ala Phe
831                      40                      45                      50
832
833 GAC ACC TAC CAG GAG TTT GAA GAA GCC TAT ATC CCA AAG GAA CAG AAG      307
834 Asp Thr Tyr Gln Glu Phe Glu Glu Ala Tyr Ile Pro Lys Glu Gln Lys
835                      55                      60                      65
836
837 TAT TCA TTC CTGCA      321
838 Tyr Ser Phe
839      70
840
841
842 (2) INFORMATION FOR SEQ ID NO:24:
843
844 (i) SEQUENCE CHARACTERISTICS:
845 (A) LENGTH: 70 amino acids
846 (B) TYPE: amino acid
847 (D) TOPOLOGY: linear
848
849 (ii) MOLECULE TYPE: protein
850
851 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:
852
853 Met Ala Thr Gly Ser Arg Thr Ser Leu Leu Leu Ala Phe Gly Leu Leu
854      1                      5                      10                      15
855
856 Cys Leu Pro Trp Leu Gln Glu Gly Ser Ala Phe Pro Thr Ile Pro Leu
857      20                      25                      30
858
859 Ser Arg Leu Phe Asp Asn Ala Met Leu Arg Ala His Arg Leu His Gln
860      35                      40                      45
861
862 Leu Ala Phe Asp Thr Tyr Gln Glu Phe Glu Glu Ala Tyr Ile Pro Lys
863      50                      55                      60
864
865 Glu Gln Lys Tyr Ser Phe
866      65                      70
867

```

RAW SEQUENCE LISTING  
PATENT APPLICATION US/07/920,519DATE: 09/22/93  
TIME: 09:28:07

INPUT SET: S532.raw

## 868 (2) INFORMATION FOR SEQ ID NO:25:

869

## 870 (i) SEQUENCE CHARACTERISTICS:

871 (A) LENGTH: 74 base pairs

872 (B) TYPE: nucleic acid

873 (C) STRANDEDNESS: double

874 (D) TOPOLOGY: linear

875

876 (ii) MOLECULE TYPE: DNA (genomic)

877

878 (iii) HYPOTHETICAL: NO

879

880

881 (vii) IMMEDIATE SOURCE:

882 (B) CLONE: ClaI-NdeI fragment

883

884

885 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

886

887 CGATAGCGTA TAATGTGTGG AATTGTGAGC GGATAACAAT TTCACACAGT TTTTCGCGAA 60

888

889 GAAGGAGATA TACA 74

890

## 891 (2) INFORMATION FOR SEQ ID NO:26:

892

## 893 (i) SEQUENCE CHARACTERISTICS:

894 (A) LENGTH: 190 base pairs

895 (B) TYPE: nucleic acid

896 (C) STRANDEDNESS: double

897 (D) TOPOLOGY: linear

898

899 (ii) MOLECULE TYPE: DNA (genomic)

900

901 (iii) HYPOTHETICAL: NO

902

903

904 (vii) IMMEDIATE SOURCE:

905 (B) CLONE: Plasmid p373,2 fragment

906

907

908 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

909

910 GATCTTCAAG CAGACCTACA GCAAGTTCGA CACAAACTCA CACAACGATG ACGCACTACT 60

911

912 CAAGAACTAC GGGCTGCTCT ACTGCTTCAG GAAGGACATG GACAAGGTCG AGACATTCTT 120

913

914 GCGCATCGTG CAGTGCCGCT CTGTGGAGGG CAGCTGTGGC TTCTAGTAAG GTACCCTGCC 180

915

916 CTACGTACCA 190

917

## 918 (2) INFORMATION FOR SEQ ID NO:27:

RAW SEQUENCE LISTING  
PATENT APPLICATION US/07/920,519DATE: 09/22/93  
TIME: 09:28:15

INPUT SET: S532.raw

919  
920 (i) SEQUENCE CHARACTERISTICS:  
921 (A) LENGTH: 48 base pairs  
922 (B) TYPE: nucleic acid  
923 (C) STRANDEDNESS: single  
924 (D) TOPOLOGY: linear  
925  
926 (ii) MOLECULE TYPE: DNA (genomic)  
927  
928 (iii) HYPOTHETICAL: NO  
929  
930  
931 (vii) IMMEDIATE SOURCE:  
932 (B) CLONE: AccI-NdeI synthetic fragment  
933  
934  
935 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:  
936  
937 TATGTCTGCG GTAAAAGCAG CGCGCTACGG CAAGGACAAT GTTCGCGT 48  
938  
939 (2) INFORMATION FOR SEQ ID NO:28:  
940  
941 (i) SEQUENCE CHARACTERISTICS:  
942 (A) LENGTH: 360 base pairs  
943 (B) TYPE: nucleic acid  
944 (C) STRANDEDNESS: single  
945 (D) TOPOLOGY: linear  
946  
947 (ii) MOLECULE TYPE: DNA (genomic)  
948  
949 (iii) HYPOTHETICAL: NO  
950  
951  
952 (vii) IMMEDIATE SOURCE:  
953 (B) CLONE: Plasmid pEMR469 fragment  
954  
955  
956 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:  
957  
958 GGGACGCGTC TCCTCTGCCG GAACACCGGG CATCTCCAAC TTATAAGTTG GAGAAATAAG 60  
959  
960 AGAATTTTCAG ATTGAGAGAA TGAAAAAAAA AAAAAAAAAA AAGGCAGAGG AGAGCATAGA 120  
961  
962 AATGGGGTTC ACTTTTGGT AAAGCTATAG CATGCCTATC ACATATAAAT AGAGTGCCAG 180  
963  
964 TAGCGACTTT TTTCACACTC GAGATACTCT TACTACTGCT CTCTTGTTGT TTTTATCACT 240  
965  
966 TCTTGTTTCT TCTTGGTAAA TAGAATATCA AGCTACAAAA AGCATACAAT CAACTATCAA 300  
967  
968 CTATTA ACTA TATCGATACC ATATGGATCC GTCGACTCTA GAGGATCGTC GACTCTAGAG 360  
969

RAW SEQUENCE LISTING  
PATENT APPLICATION US/07/920,519DATE: 09/22/93  
TIME: 09:28:23

INPUT SET: S532.raw

970  
971 (2) INFORMATION FOR SEQ ID NO:29:  
972  
973 (i) SEQUENCE CHARACTERISTICS:  
974 (A) LENGTH: 58 base pairs  
975 (B) TYPE: nucleic acid  
976 (C) STRANDEDNESS: double  
977 (D) TOPOLOGY: linear  
978  
979 (ii) MOLECULE TYPE: DNA (genomic)  
980  
981 (iii) HYPOTHETICAL: NO  
982  
983  
984 (vii) IMMEDIATE SOURCE:  
985 (B) CLONE: Fragment C  
986  
987  
988 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:  
989  
990 CGATATACAC AATGTCTGCT GTTAAGGCTG CTAGATACGG TAAGGACAAC GTTAGAGT 58  
991  
992 (2) INFORMATION FOR SEQ ID NO:30:  
993  
994 (i) SEQUENCE CHARACTERISTICS:  
995 (A) LENGTH: 1013 base pairs  
996 (B) TYPE: nucleic acid  
997 (C) STRANDEDNESS: double  
998 (D) TOPOLOGY: linear  
999  
1000 (ii) MOLECULE TYPE: DNA (genomic)  
1001  
1002 (iii) HYPOTHETICAL: NO  
1003  
1004  
1005 (vii) IMMEDIATE SOURCE:  
1006 (B) CLONE: Fragment D  
1007  
1008  
1009 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:  
1010  
1011 CTACAAGGTT CACAAGGACG AGAAGACCGG TGTCCAGACG GTGTACGAGA TGACCGTCTG 60  
1012  
1013 TGTGCTTCTG GAGGGTGAGA TTGAGACCTC TTACACCAAG GCCGACAACA GCGTCATTGT 120  
1014  
1015 CGCAACCGAC TCCATTAAGA ACACCATTTA CATCACCGCC AAGCAGAACC CCGTTACTCC 180  
1016  
1017 TCCCGAGCTG TTCGGCTCCA TCCTGGGCAC ACACTTCATT GAGAAGTACA ACCACATCCA 240  
1018  
1019 TGCCGCTCAC GTCAACATTG TCTGCCACCG CTGGACCCGG ATGGACATTG ACGGCAAGCC 300  
1020

RAW SEQUENCE LISTING  
PATENT APPLICATION US/07/920,519DATE: 09/22/93  
TIME: 09:28:31

INPUT SET: S532.raw

|      |  |      |
|------|--|------|
| 1021 | ACACCCCTCAC TCCTTCATCC GCGACAGCGA GGAGAAGCGG AATGTGCAGG TGGACGTGGT | 360  |
| 1022 |  |      |
| 1023 | CGAGGGCAAG GGCATCGATA TCAAGTCGTC TCTGTCCGGC CTGACCGTGC TGAAGAGCAC  | 420  |
| 1024 |  |      |
| 1025 | CAACTCGCAG TTCTGGGGCT TCCTGCGTGA CGAGTACACC AACTTAAGG AGACCTGGGA   | 480  |
| 1026 |  |      |
| 1027 | CCGTATCCTG AGCACCGACG TCGATGCCAC TTGGCAGTGG AAGAATTTCA GTGGACTCCA  | 540  |
| 1028 |  |      |
| 1029 | GGAGGTCCGC TCGCACGTGC CTAAGTTCGA TGCTACCTGG GCCACTGCTC GCGAGGTCAC  | 600  |
| 1030 |  |      |
| 1031 | TCTGAAGACT TTTGCTGAAG ATAACAGTGC CAGCGTGCAG GCCACTATGT ACAAGATGGC  | 660  |
| 1032 |  |      |
| 1033 | AGAGCAAATC CTGGCGCGCC AGCAGCTGAT CGAGACTGTC GAGTACTCGT TGCCTAACAA  | 720  |
| 1034 |  |      |
| 1035 | GCACTATTTT GAAATCGACC TGAGCTGGCA CAAGGGCCTC CAAAACACCG GCAAGAACGC  | 780  |
| 1036 |  |      |
| 1037 | CGAGGTCTTC GCTCCTCAGT CGGACCCCAA CGGTCTGATC AAGTGTACCG TCGGCCGGTC  | 840  |
| 1038 |  |      |
| 1039 | CTCTCTGAAG TCTAAATTGT AAACCAACAT GATTCTCAG TTCCGGAGTT TCCAAGGCAA   | 900  |
| 1040 |  |      |
| 1041 | ACTGTATATA GTCTGGGATA GGGTATAGCA TTCATTCCT TGTTTTTTAC TTCCAAAAAA   | 960  |
| 1042 |  |      |
| 1043 | AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAGGGC CCG         | 1013 |
| 1044 |  |      |
| 1045 | (2) INFORMATION FOR SEQ ID NO:31:                                  |      |
| 1046 |  |      |
| 1047 | (i) SEQUENCE CHARACTERISTICS:                                      |      |
| 1048 | (A) LENGTH: 207 base pairs   |      |
| 1049 | (B) TYPE: nucleic acid   |      |
| 1050 | (C) STRANDEDNESS: double   |      |
| 1051 | (D) TOPOLOGY: linear   |      |
| 1052 |  |      |
| 1053 | (ii) MOLECULE TYPE: DNA (genomic)                                  |      |
| 1054 |  |      |
| 1055 | (iii) HYPOTHETICAL: NO   |      |
| 1056 |  |      |
| 1057 |  |      |
| 1058 | (vii) IMMEDIATE SOURCE:  |      |
| 1059 | (B) CLONE: Synthetic GAL7 fragment                                 |      |
| 1060 |  |      |
| 1061 |  |      |
| 1062 | (xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:                           |      |
| 1063 |  |      |
| 1064 | CGCGTCTATA CTTCGGAGCA CTGTTGAGCG AAGGCTCATT AGATATATTT TCTGTCATTT  | 60   |
| 1065 |  |      |
| 1066 | TCCTTAACCC AAAAATAAGG GAGAGGGTCC AAAAAGCGCT CGGACAACCTG TTGACCGTGA | 120  |
| 1067 |  |      |
| 1068 | TCCGAAGGAC TGGCTATACA GTGTTACAAA AATAGCCAAG CTGAAAATAA TGTGTAGCCT  | 180  |
| 1069 |  |      |
| 1070 | TTAGCTATGT TCAGTTAGTT TGGCATG                                      | 207  |
| 1071 |  |      |

RAW SEQUENCE LISTING  
PATENT APPLICATION US/07/920,519DATE: 09/22/93  
TIME: 09:28:39

INPUT SET: S532.raw

1072 (2) INFORMATION FOR SEQ ID NO:32:  
1073  
1074 (i) SEQUENCE CHARACTERISTICS:  
1075 (A) LENGTH: 23 base pairs  
1076 (B) TYPE: nucleic acid  
1077 (C) STRANDEDNESS: single  
1078 (D) TOPOLOGY: linear  
1079  
1080 (ii) MOLECULE TYPE: DNA (genomic)  
1081  
1082 (iii) HYPOTHETICAL: NO  
1083  
1084  
1085 (vii) IMMEDIATE SOURCE:  
1086 (B) CLONE: Modified XbaI-MluI adapter  
1087  
1088  
1089 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:  
1090  
1091 CTAGGCTAGC GGGCCCGCAT GCA 23  
1092  
1093 (2) INFORMATION FOR SEQ ID NO:33:  
1094  
1095 (i) SEQUENCE CHARACTERISTICS:  
1096 (A) LENGTH: 422 base pairs  
1097 (B) TYPE: nucleic acid  
1098 (C) STRANDEDNESS: single  
1099 (D) TOPOLOGY: linear  
1100  
1101 (ii) MOLECULE TYPE: DNA (genomic)  
1102  
1103 (iii) HYPOTHETICAL: NO  
1104  
1105  
1106 (vii) IMMEDIATE SOURCE:  
1107 (B) CLONE: Plasmid pSE1 "site binding to HindIII"  
1108 fragment  
1109  
1110  
1111 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:  
1112  
1113 AGCTGGCTCG CATCTCTCCT TCACGCGCCC GCCGCCCTAC CTGAGGCCGC CATCCACGCC 60  
1114  
1115 GGTGAGTCGC GTTCTGCCGC CTCCCGCCTG TGGTGCTCC TGAAGTGGT CCGCCGTCTA 120  
1116  
1117 GGTAGGCTCC AAGGGAGCCG GACAAAGGCC CGGTCTCGAC CTGAGCTCTA AACTTACCTA 180  
1118  
1119 GACTCAGCCG GCTCTCCACG CTTTGCCTGA CCCTGCTTGC TCAACTCTAC GTCTTTGTTT 240  
1120  
1121 CGTTTTCTGT TCTGCGCCGT TACAACTTCA AGGTATGCGC TGGGACCTGG CAGGCGGCAT 300  
1122

RAW SEQUENCE LISTING  
PATENT APPLICATION US/07/920,519DATE: 09/22/93  
TIME: 09:28:47

INPUT SET: S532.raw

1123 CTGGGACCCC TAGGAAGGGC TTGGGGGTCC TCGTGCCCAA GGCAGGGAAC ATAGTGGTCC 360  
1124  
1125 CAGGAAGGGG AGCAGAGGCA TCAGGGTGTC CACTTTGTCT CCGCAGCTCC TGAGCCTGCA 420  
1126  
1127 GA 422  
1128  
1129 (2) INFORMATION FOR SEQ ID NO:34:  
1130  
1131 (i) SEQUENCE CHARACTERISTICS:  
1132 (A) LENGTH: 77 base pairs  
1133 (B) TYPE: nucleic acid  
1134 (C) STRANDEDNESS: double  
1135 (D) TOPOLOGY: linear  
1136  
1137 (ii) MOLECULE TYPE: DNA (genomic)  
1138  
1139 (iii) HYPOTHETICAL: NO  
1140  
1141  
1142 (vii) IMMEDIATE SOURCE:  
1143 (B) CLONE: Synthetic HindIII-"site binding to BamHI"  
1144 fragment  
1145  
1146  
1147 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:  
1148  
1149 AGCTTGTCGA CTAATACGAC TCACTATAGG GCGGCCGCGG GCCCCTGCAG GAATTCGGAT 60  
1150  
1151 CCCCCGGGTG ACTGACT 77  
1152  
1153 (2) INFORMATION FOR SEQ ID NO:35:  
1154  
1155 (i) SEQUENCE CHARACTERISTICS:  
1156 (A) LENGTH: 61 base pairs  
1157 (B) TYPE: nucleic acid  
1158 (C) STRANDEDNESS: double  
1159 (D) TOPOLOGY: linear  
1160  
1161 (ii) MOLECULE TYPE: DNA (genomic)  
1162  
1163 (iii) HYPOTHETICAL: NO  
1164  
1165  
1166 (vii) IMMEDIATE SOURCE:  
1167 (B) CLONE: Synthetic HindIII-AccI fragment  
1168  
1169  
1170 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:  
1171  
1172 AGCTTGCCGC CACTATGTCC GCAGTAAAAG CAGCCCGCTA CGGCAAGGAC AATGTCCGCG 60  
1173

RAW SEQUENCE LISTING  
PATENT APPLICATION US/07/920,519DATE: 09/22/93  
TIME: 09:28:55

INPUT SET: S532.raw

1174 T 61  
1175  
1176 (2) INFORMATION FOR SEQ ID NO:36:  
1177  
1178 (i) SEQUENCE CHARACTERISTICS:  
1179 (A) LENGTH: 920 base pairs  
1180 (B) TYPE: nucleic acid  
1181 (C) STRANDEDNESS: single  
1182 (D) TOPOLOGY: linear  
1183  
1184 (ii) MOLECULE TYPE: DNA (genomic)  
1185  
1186 (iii) HYPOTHETICAL: NO  
1187  
1188  
1189 (vii) IMMEDIATE SOURCE:  
1190 (B) CLONE: HindIII-SnaBI fragment  
1191  
1192  
1193 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:  
1194  
1195 AGCTTGCCGC CACTATGTCC GCAGTAAAAG CAGCCCGCTA CGGCAAGGAC AATGTCCGCG 60  
1196  
1197 TCTACAAGGT TCACAAGGAC GAGAAGACCG GTGTCCAGAC GGTGTACGAG ATGACCGTCT 120  
1198  
1199 GTGTGCTTCT GGAGGGTGAG ATTGAGACCT CTTACACCAA GGCCGACAAC AGCGTCATTG 180  
1200  
1201 TCGCAACCGA CTCCATTAAG AACACCATTT ACATCACCGC CAAGCAGAAC CCCGTTACTC 240  
1202  
1203 CTCCCGAGCT GTTCGGCTCC ATCCTGGGCA CACACTTCAT TGAGAAGTAC AACCACATCC 300  
1204  
1205 ATGCCGCTCA CGTCAACATT GTCTGCCACC GCTGGACCCG GATGGACATT GACGGCAAGC 360  
1206  
1207 CACACCCTCA CTCCTTCATC CGCGACAGCG AGGAGAAGCG GAATGTGCAG GTGGACGTGG 420  
1208  
1209 TCGAGGGCAA GGGCATCGAT ATCAAGTCGT CTCTGTCCGG CCTGACCGTG CTGAAGAGCA 480  
1210  
1211 CCAACTCGCA GTTCTGGGGC TTCCTGCGTG ACGAGTACAC CACACTTAAG GAGACCTGGG 540  
1212  
1213 ACCGTATCCT GAGCACCGAC GTCGATGCCA CTTGGCAGTG GAAGAATTTC AGTGGACTCC 600  
1214  
1215 AGGAGGTCCG CTCGCACGTG CCTAAGTTCTG ATGCTACCTG GGCCACTGCT CGCGAGGTCA 660  
1216  
1217 CTCTGAAGAC TTTTGCTGAA GATAACAGTG CCAGCGTGCA GGCCACTATG TACAAGATGG 720  
1218  
1219 CAGAGCAAAT CCTGGCGCGC CAGCAGCTGA TCGAGACTGT CGAGTACTCG TTGCCTAACA 780  
1220  
1221 AGCACTATTT CGAAATCGAC CTGAGCTGGC ACAAGGGCCT CCAAAACACC GGCAAGAACG 840  
1222  
1223 CCGAGGTCTT CGCTCCTCAG TCGGACCCCA ACGGTCTGAT CAAGTGTACC GTCGGCCGGT 900  
1224

PAGE: 25

**RAW SEQUENCE LISTING**  
**PATENT APPLICATION US/07/920,519**

DATE: 09/22/93  
TIME: 09:29:03

*INPUT SET: S532.raw*

1225 CCTCTCTGAA GTCTAAATTG  
1226

920

PAGE: 1

**SEQUENCE VERIFICATION REPORT**  
**PATENT APPLICATION US/07/920,519**

DATE: 09/22/93  
TIME: 09:29:03

*INPUT SET: S532.raw*

| Line | Error                           | Original Text                         |
|------|---------------------------------|---------------------------------------|
| 37   | Wrong application Serial Number | (A) APPLICATION NUMBER: US/07/920,519 |

PAGE: 1

**SEQUENCE MISSING ITEM REPORT**  
**PATENT APPLICATION *US/07/920,519***

DATE: 09/22/93  
TIME: 09:29:03

*INPUT SET: S532.raw*

< < THERE ARE NO ITEMS MISSING > >

PAGE: 1

**SEQUENCE CORRECTION REPORT**  
**PATENT APPLICATION US/07/920,519**

DATE: 09/22/93  
TIME: 09:29:04

*INPUT SET: S532.raw*

| Line | Original Text | Corrected Text |
|------|---------------|----------------|
|------|---------------|----------------|